

DJ ミックス品質の測定可能性について

トランジション評価への信号ベースアプローチ

Daito Manabe (真鍋大度)

DJ Mix Research Institute

東京, 日本

2026 年 3 月

要旨

DJ トランジションの質は、しばしばセンス、勘、経験といった語で説明される。この言い方は一面では正しいが、一面では問題を曖昧にする。音楽的判断、選曲、フロアの読みが単一のスカラー量に還元できないという意味では正しい。しかし、二つのステレオデッキ信号とステレオのマスター出力を同時に観測できるなら、トランジション品質のかなり大きな部分は客観的に測定可能になるという事実を覆い隠してしまう。本稿は、DJ トランジション評価を、二つのステレオ入力 A , B と一つのステレオ出力 C を持つ参照付き信号解析問題として定式化する。提案枠組みはまず、 A と B から C への時間依存かつ帯域依存の寄与を推定し、その寄与に基づいて、単純なエネルギー閾値ではなくトランジション区間を検出する。その上で、ラウドネス制御と true peak ヘッドルーム、スペクトル衝突、スペクトル連続性、ゲイン軌道の滑らかさ、ステレオ場の安定性、さらに任意成分としてビート位相整合性のスコアを導出する。また、単純な加法モデルが観測されたマスター信号をどの程度うまく説明しているかを表す信頼度スコアも導入する。本システムは、センス、物語性、あるいは観客反応までを測るとは主張しない。その代わりに、チュータリングソフトウェア、補助的 DJ ツール、オフライン性能解析を支えられるだけの安定性を持つ、トランジションの工学的側面と知覚的連続性の側面を切り出す。最後に、提案スコアが満たすべき基本性質、実装上の推奨デフォルト、そして専門家評価に対するキャリブレーション手順を述べる。

キーワード: DJ 解析, トランジション評価, 音響信号処理, 音楽情報検索, ラウドネス, スペクトル衝突, ビート整合性.

1 はじめに

DJ は単にトラックを順番に再生する存在ではない。その技術の中核にあるのはトランジションである。すなわち、セットのエネルギー、バランス、方向性を崩壊させることなく、一方の曲から他方の曲へ空間を受け渡す制御された時間区間である。現実には、DJ は反復と批評を通じてこれを学ぶ。典型的なフィードバックは定性的で曖昧だ。ローが喧嘩した、ブレンドが大きくなり過ぎた、切り替えが急すぎた、トランジションがビートからずれた、ステレオ像が散らかった、といった言い方である。こうした判断は有用ではあるが、仕様にはなっていない。トレーニングソフトウェア、解析ツール、補助システムを作るなら、このフィードバックは形式化されなければならない。

第一の誤りは、DJ の上手さのすべてが測定できると考えることだ。そうではない。選曲、セット全体のアーク、文化的文脈、その場のタイミング、意図的なルール破りの一部は、信号解析の

外側にある。第二の誤りはその逆で、芸術が完全には測定できない以上、意味のあるものは何も測れないと考えることだ。これは誤りである。二つのデッキ信号とマスター出力が利用できるなら、トランジション品質のうち大きく、しかも運用上重要な部分は十分に扱える。そこでは、両曲が本当にマスターに寄与しているか、マスターが不必要に大きくなっていないか、ロー帯域が衝突していないか、音色バランスが唐突に変化していないか、クロスフェード軌道が荒れていないか、ビート位相が保たれているかを定量化できる。

本稿は、この主張を具体的な枠組みに落とし込む。二つのステレオ入力信号 $A[n]$, $B[n]$ と、一つのステレオ出力信号 $C[n]$ を考える。ここで A と B はミキサーに与えられる二つのチャンネル入力あるいはデッキ出力であり、 C は観測されたステレオのマスター出力である。課題は、トランジション区間を検出し、その品質を採点することである。提案法は *reference aware* である。すなわち、ブラインドな DJ ミックストランスクリプションとは異なり、二つの寄与源が観測されているという事実を直接利用する。この点により、問題はより制約され、リアルタイムのチュータリングやスタジオ解析に向けたものになる。

設計原理は三つある。

第一に、トランジション検出は生のエネルギーではなく寄与に基づくべきである。 A と B に単純な閾値を置くやり方は、静かなパッセージ、ブレイク、cue bleed、あるいは両入力がある場合でも片方が実質的にはマスターに出ていない場面で誤動作する。第二に、評価は多成分であるべきである。内訳のない単一スコアは、何が悪かったかを説明できないので、ほとんど役に立たない。第三に、システムは適用範囲に正直であるべきである。測定可能な部分だけを採点し、加法的信号モデルでマスターがうまく説明できないときは信頼度を下げて返すべきである。

したがって本稿の貢献は四つある。第一に、トランジション評価を、二つの観測されたステレオ入力と一つの観測されたステレオ出力の下での信号解析問題として定式化する。第二に、平滑な非負スペクトル回帰に基づく、寄与認識型のトランジション検出器を提案する。第三に、ラウドネス、スペクトル衝突、連続性、滑らかさ、ステレオ安定性、ビート整合性の成分スコアを導出する。第四に、実装を可能にするための基本性質、推奨デフォルト、キャリブレーション手順を与える。

2 関連研究

2.1 自動 DJ とトランジション生成

自動 DJ は二十年以上にわたって研究されてきた。Cliff による *Hang the DJ* と *hpDJ* は、シーケンシングとシームレスなミキシングを、トラック適合性、トランジション計画、ミックススタイルのユーザ制御を含む計算問題として定式化した [1, 2]。Ishizaki らは、ユーザ不快感関数に基づくテンポ調整を備えた全自動 DJ ミキシングシステムを提案した [3]。Hirai らの *MusicMixer* は、ビート構造と潜在トピック類似度によりトランジション選択を行った [4]。Bittner らは、より大規模な楽曲群に対してプレイリストの順序付けとクロスフェード遷移設計を定式化した [5]。その後、Vande Veire と De Bie は、ドラムベース向けの包括的な自動 DJ パイプラインを提示した [6]。さらに近年では、実際のミックスから学習したイコライザとフェーダの明示的モジュールを用いる微分可能なトランジション合成も報告されている [7]。Williams らによる最新の概観は、この分野をトラックレベル、トランジションレベル、ミックスレベルという時間スケールで整理している [8]。一方、ハーモニックミキシングの系譜は、既に実行されたトランジションの品質よりも、候補曲同士の調性的適合性を扱う [9]。

2.2 DJ ミックス解析, リバースエンジニアリング, データセット

第二の研究群は、録音された DJ ミックスを生成対象ではなく測定対象として扱う。Schwarz と Fourer は DJ ミックスからの ground truth 抽出を開始し [10], DJ ミックス情報検索のための UnmixDB データセットを公開し [11], さらに DJ ミックス逆解析の手法とデータセットを整理した [12]。Kim らは、mix-to-track subsequence alignment によって実世界の DJ ミックスを大規模に解析できることを示した [13]。続く研究では、同系統の方法が凸最適化によりトランジション区間を逆解析し [14], 大規模ミックス群からフェーダゲインとイコライザゲインを同時推定している [15]。最近では、André らが multi-pass 非負値行列因子分解により blind DJ mix transcription をさらに前進させた [16]。これらの研究は、意味のあるミキサ挙動が音響から回復できることを示しているが、主眼は再構成, 転写, データセット構築にあり, 解釈可能なトランジション品質スコアリングそのものではない。

2.3 キューポイントとトランジション境界推定

関連研究は、トランジションがどこで起こりうるか、あるいは起こるべきかも扱っている。Schwarz, Schindler, Spadavecchia は DJ キューポイント推定のヒューリスティックアルゴリズムを提案した [17]。Zehren, Alunno, Bientinesi は M-DJCUE データセットを導入し [18], さらに DJ ミキシングのエミュレーションのための自動キューポイント検出を開発した [19]。Argüello, Lanzendörfer, Wattenhofer は最近、キューポイント推定を物体検出問題として再定式化し、およそ 5,000 曲に対して約 21,000 個の専門家アノテーション済みキューポイントを含む、より大規模なデータセットを公開した [20]。キューポイント研究は本稿に隣接しているが、同一ではない。それらが扱うのは妥当なトランジションのアンカーや構造境界であり、本稿が扱うのは既に起こったトランジションの実行品質である。

2.4 インテリジェント音楽ミキシングと知覚評価

DJ の外側では、インテリジェント音楽制作研究が、バランスと明瞭性を測定可能な工学量として扱うための最も強い先例を提供する。初期のルールベース自動ミキシング研究は、パンニング, ゲイン制御, ライブコンソール挙動を扱った [21, 22, 23]。Pestana と Reiss は、ミキシングのベストプラクティスに基づく自動制作戦略を整理した [24]。De Man らはマルチトラック混合信号の音響特徴を分析し [25], 音楽ミキシング実践の知覚評価を行い [26], さらに十年分の自動ミキシング研究を総括した [27]。より広い intelligent music production のレビューは、これらを解析, 変換, アシスタントシステムの大きな文脈に位置付ける [28]。近年の学習ベース手法には、Wave-U-Net を用いたインテリジェントドラムミキシング [29] や、経験豊富なミキシングエンジニアによって検証された out-of-domain データによる自動音楽ミキシング [30] がある。この文献群は、ラウドネス, スペクトルバランス, 空間配置を測定対象とみなす具体的な前例を与えるが、解析単位は通常スタジオのマルチトラックミックスであり、二曲間のライブ DJ トランジションではない。

2.5 参照付きミックス逆推定と客観評価

本稿の設定に最も近い方法論上の先行研究は、参照付きミックス逆推定である。すなわち、ソース信号と結果としてのミックスが同時に観測されるときに、ミキサ挙動や信号処理鎖を推定する枠組みである。Barchiesi と Reiss は、マルチトラック録音と目標ステレオミックスからミックスを逆解析した [31]。Ramona と Richard は、多入力一出力の放送信号からコンソールフェーダ位置

表 1: 本稿と隣接研究領域との位置関係.

研究テーマ	代表的研究	本稿との関係
自動 DJ とトランジション生成	Cliff; Ishizaki ら; Hirai ら; Bittner ら; Vande Veire and De Bie; Chen ら [1, 3, 4, 5, 6, 7]	トラック選択やトランジション合成を主題としており, 観測済みトランジションの実行品質採点は主眼ではない.
DJ ミックス MIR, 逆解析, データセット	Schwarz and Fourer; Kim ら; André ら [10, 11, 12, 13, 14, 15, 16]	DJ 固有の最も近い先行系譜である. 主眼は再構成, 転写, データセット構築であり, 解釈可能な品質スコアではない.
キューポイントとトランジション境界	Schwarz ら; Zehren ら; Argüello ら [17, 18, 19, 20]	トランジションがどこで起こるべきかを扱う. 本稿は既に起きたトランジションがどれだけうまく実行されたかを評価する.
インテリジェントミキシングと知覚研究	Perez-Gonzalez and Reiss; Pestana and Reiss; De Man ら; Moffat and Sandler; Martínez-Ramírez ら [21, 22, 23, 24, 25, 26, 27, 28, 29, 30]	バランスに関する測定概念と評価実践を与えるが, 通常は DJ トランジションではなくスタジオのマルチトラックミキシングを対象とする.
参照付きミックス逆推定	Barchiesi and Reiss; Ramona and Richard; Colonel and Reiss; Lee ら [31, 32, 33, 34]	観測された入力と出力を共同利用するという中核発想を与える. 本稿はそれをトランジション検出と品質スコアリングへ適用する.

を推定した [32]. Colonel と Reiss はその後, 微分可能 DSP を用いて raw tracks とステレオミックスダウンからミックスパラメータを回復した [33]. 最近の研究はこれをさらに進め, グラフ構造を持つミキシングチェーンの逆解析へ拡張している [34]. 評価側では, ITU-R BS.1770 のようなラウドネス標準が, 番組ラウドネスと true peak の安定な運用指標を提供する [35]. また, ビートトラッキングとリアルタイムリズム解析は, リズム依存トランジション評価に使える位相特徴を提供する [36, 37, 38]. 同時に, 客観的音質指標のレビューは, 指標の性能が適用ドメインに強く依存することを警告している [39]. この警告は本稿にとって重要である. DJ トランジションのスコアは, トランジション特有の失敗モードを考えずに任意の音質指標を持ち込んで組み立てるべきではない.

本稿はこれらの文献群の中間に位置する. blind DJ mix transcription よりも情報量の多い観測モデルを仮定し, 二つのデッキ信号とマスター信号の両方を利用する. 一方で, 自動トランジション合成が問う「どのトランジションを生成すべきか」とは異なる問いを立てる. すなわち, 既に実行されたトランジションが, ラウドネス, スペクトル占有, 連続性, ステレオ挙動, リズム整合をどれだけ適切に制御できたかである. 私たちの知る限り, 先行研究には, 二つの観測されたステレオ入力と一つの観測されたステレオ出力に対して, これらの次元を参照付きスコアリング枠組みに統合したものはない.

3 問題設定と適用範囲

二つのステレオ入力信号と一つのステレオ出力信号を観測する.

$$A[n], B[n], C[n] \in \mathbb{R}^2, \quad n = 0, \dots, N-1, \quad (1)$$

ここで A と B は二つのチャンネル入力, C はステレオのマスター出力を表す。システムの出力は単なるスカラーではなく, 次のような構造化レポートである。

$$\mathcal{R}(T) = (Q, \mathbf{S}, \text{Conf}, T), \quad (2)$$

ここで $T = [m_s, m_e]$ はフレーム単位で検出されたトランジション区間, \mathbf{S} は成分スコアのベクトル, Q は 0 から 100 の任意総合スコア, Conf は採用した信号モデルの下でそのスコアがどれだけ信用できるかを表す。

適用範囲は意図的に狭くとる。私たちは, 選曲同士が美学的に合っているか, セット順序に物語的強さがあるか, あるいは特定の観客に対して DJ が適切なフロアレベル判断をしたかまでは測ろうとしない。一方で, トランジション区間が, 聴感上も技術的にも重要な次の問いに対してどれだけ制御されているかは測ろうとする。

1. 主張されたトランジション中に, 両トラックは本当にマスターへ現れているか。
2. 重なりは回避可能なラウドネス過多や true peak 問題を生んでいないか。
3. 二つのトラックは, 特に低域において同じ周波数空間を奪い合っていないか。
4. スペクトル包絡は, 飛び跳ねるのではなく一貫して推移しているか。
5. スタイルが意図的なカットを要求しない限り, ゲイン軌道は滑らかか。
6. ステレオ場は安定しているか。
7. リズムが重要なとき, ビート位相は整合しているか。

この範囲だけでも十分に有用である。これらの問いに答えられないチュータリングシステムは, ほとんど装飾品にすぎない。

4 寄与認識型トランジション検出

4.1 時間周波数表現

各観測ステレオ信号 $X \in \{A, B, C\}$ に対し, mid と side チャンネルを

$$X^M = \frac{X_L + X_R}{\sqrt{2}}, \quad X^S = \frac{X_L - X_R}{\sqrt{2}}, \quad (3)$$

として構成し, フレーム m と周波数ビン k に対して重なりを持つ short time Fourier transform を計算する。mid エネルギーはラウドネスとスペクトル解析に用い, side エネルギーはステレオ解析のために保持する。さらに, スペクトルを B 個の聴覚帯域 $\{B_b\}_{b=1}^B$ に集約し, 帯域エネルギーを

$$e_X(m, b) = \sum_{k \in B_b} |X_m(k)|^2. \quad (4)$$

で定義する。

実装では, 44.1 または 48 kHz において, 2048 あるいは 4096 点窓, 512 から 1024 サンプル程度の hop を用いればよい。重要なのはフロントエンドの細部よりも, 三信号間で一貫性があることである。

4.2 平滑な非負寄与推定

「両デッキが大きければトランジションがある」という典型的な閾値規則は、土台として使えないほど頻繁に外れる。重要なのは、両デッキがマスターに寄与しているかである。そこで、各フレーム、各帯域において、平滑な非負振幅モデルを当てはめることで寄与を推定する。

$$(\hat{g}_A(m, b), \hat{g}_B(m, b)) = \arg \min_{g_A, g_B \geq 0} \sum_{k \in \mathcal{B}_b} \left(|C_m(k)| - g_A |A_m(k)| - g_B |B_m(k)| \right)^2 + \lambda \left\| \begin{bmatrix} g_A \\ g_B \end{bmatrix} - \begin{bmatrix} \hat{g}_A(m-1, b) \\ \hat{g}_B(m-1, b) \end{bmatrix} \right\|_2^2. \quad (5)$$

第一項は、与えられた帯域における観測マスター振幅を説明する。第二項は、推定ゲインがフレーム間で物理的に不自然なジャンプをすることを抑制する。これはDJミキサーの完全なモデルではない。位相干渉やエフェクトのテイルは無視している。それでも、ソースが実質的に存在しているかどうかを見積もるには十分である。

推定ゲインから寄与比を

$$r_A(m, b) = \frac{\hat{g}_A(m, b)e_A(m, b)}{\hat{g}_A(m, b)e_A(m, b) + \hat{g}_B(m, b)e_B(m, b) + \varepsilon}, \quad (6)$$

$$r_B(m, b) = \frac{\hat{g}_B(m, b)e_B(m, b)}{\hat{g}_A(m, b)e_A(m, b) + \hat{g}_B(m, b)e_B(m, b) + \varepsilon}. \quad (7)$$

で定義する。さらに、 $\pi_b \geq 0$, $\sum_b \pi_b = 1$ を満たす帯域重みを用いて、フレームレベル寄与

$$\bar{r}_A(m) = \sum_{b=1}^B \pi_b r_A(m, b), \quad \bar{r}_B(m) = \sum_{b=1}^B \pi_b r_B(m, b). \quad (8)$$

を得る。するとトランジション活動度は

$$\mu(m) = \min(\bar{r}_A(m), \bar{r}_B(m)). \quad (9)$$

で与えられる。

4.3 トランジション開始, 終了, 信頼度

活動度閾値 $\tau_{\text{on}} > \tau_{\text{off}}$ と、フレーム単位の最小連続長 L_{on} , L_{off} を定める。開始フレーム m_s は、 $\mu(m) > \tau_{\text{on}}$ を満たし、かつ A と B の両方が少なくとも L_{on} フレーム連続してノイズフロアを超えた最初のフレームとする。終了フレーム m_e は、その後で $\mu(m) < \tau_{\text{off}}$ が少なくとも L_{off} フレーム続く最初のフレーム、あるいは一方の寄与が持続的に消失した時点とする。

検出器は不確実性も認めるべきである。そこで、観測マスター振幅と推定された加法説明との間の残差を

$$\rho(m, b) = \frac{\sum_{k \in \mathcal{B}_b} \left| |C_m(k)| - \hat{g}_A(m, b) |A_m(k)| - \hat{g}_B(m, b) |B_m(k)| \right|}{\sum_{k \in \mathcal{B}_b} |C_m(k)| + \varepsilon}. \quad (10)$$

で定義する。トランジション信頼度は

$$\text{Conf} = \exp \left(-\frac{1}{s_\rho |T| B} \sum_{m \in T} \sum_{b=1}^B \rho(m, b) \right), \quad (11)$$

とする。強い delay feedback, reverb bloom, クリッピング, あるいは位相依存の強いエフェクトは, たとえミックスが音楽的に成立していても, 正当に信頼度を下げうる。これは欠陥ではなく, むしろ望ましい挙動である。不確実性を伴わないスコアは虚偽に近い。

A と B がライブのチャンネル入力ではなく元のトラックファイルである場合でも, 本枠組み自体は適用できる。ただしその場合は, 事前にアライメントと変換の段階が必要になる。そこでは, mix-to-track alignment, トランジション逆解析, DJ mix transcription の文献が直接に関係する [13, 14, 15, 16]。

5 トランジション品質指標

検出されたトランジションを $T = [m_s, m_e]$ とする。本枠組みは, それぞれ単独でも解釈可能な成分スコア $S_i \in [0, 1]$ を返す。値が高いほど良く, 低いほど特定の失敗モードを示す。

5.1 ラウドネス制御とヘッドルーム

最もありふれた失敗は微妙でも何でもない。単に重なりが大きくなり過ぎることである。ここでは ITU BS.1770 系の実務に整合する short term loudness と true peak を採用する [35]。フレーム m における信号 X の short term loudness を $L_X(m)$, true peak を $P_X(m)$ とする。すると

$$\Delta_L(m) = L_C(m) - \max(L_A(m), L_B(m)). \quad (12)$$

と定義できる。LOUDNESSペナルティは

$$p_{ld} = \frac{1}{|T|} \sum_{m \in T} ([\Delta_L(m) - \delta_L]_+ + \gamma [P_C(m) - P_{\max}]_+), \quad (13)$$

であり, ここで δ_L はLOUDNESS許容値, P_{\max} はたとえば -1 dBTP のようなヘッドルーム目標である。対応するスコアは

$$S_{ld} = \exp\left(-\frac{p_{ld}}{s_{ld}}\right). \quad (14)$$

とする。この成分は意図的に厳しく設計している。スタイル上の議論に勝っても, マスターをクリップさせるトランジションは, 実用上よいトランジションではない。

5.2 スペクトル衝突

次の失敗モードは周波数競合である。二つのトラックはエネルギー的には同時に存在していても, 同じ帯域を同じ時間に占有すると, 聴感上は破壊的に干渉しうる。最も典型的なのはキックとベースの重なりだが, フック, ベース倍音, ボーカル基音が集まる lower mid でも同じことが起こる。

正規化帯域占有率を

$$q_X(m, b) = \frac{e_X(m, b)}{\sum_{b'=1}^B e_X(m, b') + \varepsilon}. \quad (15)$$

で定義する。衝突ペナルティは

$$p_{col} = \frac{1}{|T|} \sum_{m \in T} \sum_{b=1}^B \omega_b r_A(m, b) r_B(m, b) \min(q_A(m, b), q_B(m, b)), \quad (16)$$

で与える。ここで ω_b は通常、低域と lower mid を強調する知覚重みである。スコアは

$$S_{\text{col}} = \exp\left(-\frac{p_{\text{col}}}{s_{\text{col}}}\right). \quad (17)$$

とする。これは普遍的な心理音響マスキングモデルではない。あくまで重なりリスクに焦点を当てた標的指標である。

5.3 スペクトル連続性

クリーンなトランジションは、単に濁っていないだけではない。送り出す曲から受け入れる曲へ向かって、首尾一貫して移動する。マスターのスペクトルが、どちらの曲からも予測できない形で飛ぶと、耳はそれをブレンドではなく粗いスイッチとして知覚しやすい。

W_A をトランジション前で A が優勢かつ B が非活性な短い窓、 W_B をトランジション後で B が優勢かつ A が非活性な短い窓とする。log 帯域テンプレートを

$$u_A(b) = \text{median}_{m \in W_A} \log(e_A(m, b) + \varepsilon), \quad (18)$$

$$u_B(b) = \text{median}_{m \in W_B} \log(e_B(m, b) + \varepsilon). \quad (19)$$

で定義する。 $m \in T$ に対し、

$$\alpha(m) = \frac{m - m_s}{m_e - m_s + \varepsilon}, \quad (20)$$

$$u^*(m, b) = (1 - \alpha(m))u_A(b) + \alpha(m)u_B(b). \quad (21)$$

と置く。すると連続性ペナルティは

$$p_{\text{cty}} = \frac{1}{|T|B} \sum_{m \in T} \sum_{b=1}^B \nu_b |\log(e_C(m, b) + \varepsilon) - u^*(m, b)|, \quad (22)$$

であり、スコアは

$$S_{\text{cty}} = \exp\left(-\frac{p_{\text{cty}}}{s_{\text{cty}}}\right). \quad (23)$$

となる。前後のクリーンな窓が取れない場合は、同時刻のソース包絡や寄与比から期待軌道を構成してもよい。原理は変わらない。よいトランジションは移動するが、跳ねない。

5.4 ゲイン軌道の滑らかさ

多くのトランジションは一点の決断ではなく、軌道である。ビートマッチやスペクトル制御が妥当でも、軌道が荒いと急に素人っぽく聞こえる。帯域集約ゲインを

$$g_A(m) = \sum_{b=1}^B \pi_b \hat{g}_A(m, b), \quad g_B(m) = \sum_{b=1}^B \pi_b \hat{g}_B(m, b). \quad (24)$$

とする。二階差分

$$\Delta^2 g(m) = g(m+1) - 2g(m) + g(m-1), \quad (25)$$

を用いて、滑らかさペナルティを

$$p_{\text{sm}} = \frac{1}{|T| - 2} \sum_{m=m_s+1}^{m_e-1} (|\Delta^2 g_A(m)| + |\Delta^2 g_B(m)| + \eta |L_C(m+1) - L_C(m)|). \quad (26)$$

と定義する。スコアは

$$S_{\text{sm}} = \exp\left(-\frac{p_{\text{sm}}}{s_{\text{sm}}}\right). \quad (27)$$

である。カット、トランスフォーム、ハードドロップを意図的に多用するスタイルでは、この成分は重みを下げるべきである。

5.5 ステレオ安定性

トランジションは空間的にも失敗しうる。マスター像がモノラルに潰れたり、不規則に広がったり、どちらのデッキからも説明できない方向へ動いたりすることがある。mid と side のエネルギー

$$E_X^M(m) = \sum_k |X_m^M(k)|^2, \quad E_X^S(m) = \sum_k |X_m^S(k)|^2, \quad (28)$$

を用いて、ステレオ比を

$$\rho_X(m) = 10 \log_{10} \frac{E_X^M(m) + \varepsilon}{E_X^S(m) + \varepsilon}. \quad (29)$$

と定義する。期待ステレオ比は推定寄与から

$$\rho^*(m) = \bar{r}_A(m) \rho_A(m) + \bar{r}_B(m) \rho_B(m). \quad (30)$$

で与える。するとペナルティとスコアは

$$p_{\text{st}} = \frac{1}{|T|} \sum_{m \in T} |\rho_C(m) - \rho^*(m)|, \quad (31)$$

$$S_{\text{st}} = \exp\left(-\frac{p_{\text{st}}}{s_{\text{st}}}\right). \quad (32)$$

となる。位相認識型実装では inter channel coherence を加えることもできるが、mid side 比だけでも多くの明白な失敗を拾える。

5.6 ビート整合性

リズム前景のジャンルでは、位相ずれは多くの音色的誤差より速くトランジションを破壊する。任意の信頼できるビートトラックから得られる beat phase を $\phi_A(m)$ と $\phi_B(m)$ とし [36, 37, 38], 信頼性の低いフレームを抑制するリズム顕著度重みを $w(m)$ とする。ビートペナルティは

$$p_{\text{beat}} = \frac{\sum_{m \in T} w(m) \frac{1 - \cos(\phi_A(m) - \phi_B(m))}{2}}{\sum_{m \in T} w(m) + \varepsilon}, \quad (33)$$

であり、スコアは

$$S_{\text{beat}} = 1 - p_{\text{beat}}. \quad (34)$$

とする。この成分は任意である。ビートレス, free tempo, あるいは強いシンクペーションでトラック自体が信用できない場合には、省略するか大きく重みを落とすべきである。

5.7 総合スコア

最終スコアは、有効成分の重み付き平均で与える。

$$Q = 100 \frac{\sum_i \lambda_i S_i}{\sum_i \lambda_i}, \quad (35)$$

ここで $i \in \{\text{ld, col, cty, sm, st, beat}\}$ である。システムは常に Q とともに成分スコアも返すべきである。内訳のない単一数値は迷信を助長する。

表 2: 成分スコアの要約。提示する重みはあくまで開始点であり、対象ジャンルと用途に応じて専門家評価に基づき再調整すべきである。

成分	主たる手掛かり	罰するもの	推奨重み
ラウドネス制御	short term loudness と true peak	不必要に大きくなる重なり、またはヘッドルームのクリップ	0.25
スペクトル衝突	両曲が有効なときの共有帯域占有	明瞭性を覆う低域および lower mid の競合	0.25
スペクトル連続性	滑らかな source-to-source スペクトル軌道からの逸脱	粗い切り替えに聞こえる急激な音色ジャンプ	0.20
トランジション滑らかさ	ゲイン曲線の二階差分とラウドネス傾き	荒い、あるいは落ち着きのないクロスフェード挙動	0.15
ステレオ安定性	期待像からの mid side 比逸脱	説明不能なマスター像の潰れや過度な拡大	0.10
ビート整合性	円環的ビート位相差	リズム依存ジャンルでの off beat 重なり	0.05

6 妥当性の基本性質

有用なスコアは、統計的に調整される前の段階で既に妥当な振る舞いを示すべきである。以下の命題は深遠な結果ではないが重要である。これらを満たさない指標は、設計が悪い。

命題 1 (重なりゲイン追加に対する単調性). T 上の $L_A(m)$, $L_B(m)$, $P_C(m)$ を固定する。各トランジションフレームでマスターラウドネスにオフセット $\eta \geq 0$ を加え、 $L_C^\eta(m) = L_C(m) + \eta$ とする。このとき、 p_{ld} は η に対して非減少であり、 S_{ld} は η に対して非増加である。

証明. p_{ld} の各項はヒンジ関数 $[\Delta_L(m) - \delta_L]_+$ を含み、 $\Delta_L(m)$ は η に対して線形に増加する。ヒンジ関数は単調非減少なので、その平均も単調非減少である。指数写像 $x \mapsto e^{-x/s_{\text{ld}}}$ は $s_{\text{ld}} > 0$ のとき単調減少だから、スコアは非増加である。いずれかのフレームで許容値を超えれば、減少は厳密になる。□

命題 2 (共有占有率増加に対する単調性). 寄与比 $r_A(m, b)$ と $r_B(m, b)$, および全帯域重みを固定する。その他の項を固定したまま、任意のフレーム帯域対に対して $\min(q_A(m, b), q_B(m, b))$ が増加すれば、 p_{col} は減少せず、 S_{col} は増加しない。

証明. 衝突ペナルティは、非負重みと $\min(q_A, q_B)$ の積の総和である。したがって、それらの項を増やしても総和は減らない。ここでも指数写像により、非減少ペナルティは非増加スコアへ変換される。□

命題 3 (アフィンゲイン曲線は曲率ペナルティを最小化する). 離散区間上のゲイン軌道 $g(m)$ について端点を固定する。その端点を共有する全ての軌道の中で、任意のアフィン軌道は二階差分が全点でゼロであり、したがって純曲率項 $\sum_m |\Delta^2 g(m)|$ を最小化する。

証明. アフィン軌道 $g(m) = am + b$ では、二階差分は恒等的にゼロである。ペナルティは絶対値和なので、達成可能な最小値はゼロである。ゆえに任意のアフィン軌道は最小化解である。急激なステップ、折れ曲がり、振動はいずれも非ゼロの二階差分を生み、より大きなペナルティを与える。

これらの性質は、有能な DJ が直観的に既に知っていることを形式化している。重なりを押し込み過ぎれば、トランジションは悪くなる。両曲が同じ低域空間を占めれば、トランジションは悪くなる。ゲイン曲線が理由もなく暴れば、トランジションは悪くなる。

7 キャリブレーションと評価プロトコル

スコアリング枠組みは、人間の判断に対してキャリブレーションされて初めて役に立つ。問うべきなのは、ある指標が哲学的に客観的かどうかではない。専門家が妥当だと感じる順序付けを、実用に足る頻度で実現できるかどうかである。

実用的なキャリブレーション手順では、理想的には 8 秒から 20 秒程度の短いトランジション断片を、複数のジャンルとトランジションスタイルからサンプリングすべきである。各断片には、孤立した特徴量代理ではなくシステム全体を評価できるよう、観測された A, B, C 信号を付随させるべきである。その上で、専門 DJ あるいはミキシングエンジニアに対し、少なくとも四つの軸、すなわち総合的なトランジション品質、ラウドネス制御、低域のクリーンさ、ブレンドの滑らかさで評価してもらう。尺度ドリフトを減らせるので、自由数値評価より pairwise comparison の方が通常は望ましい。

キャリブレーションは三段階で進める。第一に、ラウドネス、衝突、連続性、滑らかさ、ステレオ、ビートの各スコアについてスケールパラメータを合わせ、成分スコア分布を数値的に安定させる。第二に、専門家判断に対する順位相関または pairwise accuracy を最大化するよう、成分重み λ_i を学習または調整する。第三に、ジャンル、トランジション長、ハードウェア条件をまたいだ一般化性能を検証する。統計指標としては、Spearman の順位相関、Kendall の tau, pairwise preference accuracy を含めるべきである。絶対誤差よりも順序の信頼性の方が重要だ。

自動音楽ミキシング研究で使われてきたリスニングテスト設計は、ここでも参考になる。特に casual listener ではなく経験豊富なエンジニアを採用する研究が重要である [26, 30]。ただしラベルはもっと厳密でなければならない。「production value」は DJ トランジションチュータには広すぎる。ツールが必要とするのは、DJ が実際に行動へ結びつけられる判断ラベルである。

ライブ運用では、スコアは時間方向に平滑化し、慎重に表示すべきである。チュータリング UI は単一数値を点減させるのではなく、成分トレースと説明を重視すべきだ。目的は診断である。重なり最初の四小節で collision 成分が崩壊したなら、DJ は低域が問題だと視覚的に理解できるべきである。巨大な delay wash をモデルが説明できず信頼度が下がったなら、インタフェースは確信したふりをせず、そのままそう表示すべきである。

8 限界

この枠組みには明確な限界があり、それを誤魔化すほど設計は悪くなる。

第一に、本枠組みが測るのはトランジションの実行であって、セット全体の知性ではない。技術的にはクリーンでも、悪い選曲同士をつないだトランジションは芸術的には悪い判断である。第二に、式(5)の加法振幅モデルは意図的に単純である。時間変動の強いエフェクト、非線形飽和、強い位相操作は、音楽的には説得力があるミックスであっても信頼度を下げうる。第三に、ジャンル依存性がある。緩やかなイコライゼーションを伴う長いブレンドは house や techno の一部では一般的だが、別のスタイルでは素早いカット、トランスフォーム、ドロップスワップが普通である。同じ smoothness prior が全スタイルを支配してはならない。

第四に、ビート整合性はビートトラックが正しい範囲でしか信用できない。broken rhythm, ambient,あるいはテンポ流動的な文脈では、この成分は無効化するか、大幅に重みを下げるべきである。第五に、「よいバランス」の一部は固定閾値では捉えにくいほど文脈依存である。ピークタイムのクラブセットは、自宅鑑賞向け録音ミックスよりも攻撃的な重なりを許容し、むしろ報いることすらある。

最後に、本システムは A , B , C が互換的な形式で利用できることを仮定する。利用可能な A と B がライブチャンネル信号ではなく元トラックファイルである場合、評価前に時間、テンポ、ピッチを揃えるための相当な前処理が必要である。前述の文献は有力な出発点を与えるが、問題は実質的に難しくなる [13, 16]。

これらの限界は、本枠組みを無効にするものではない。測定可能な構造と神話化された説明との境界を明示しているだけである。

9 結論

本稿は、明らかであるにもかかわらず、ほとんど形式化されてこなかった立場を主張した。DJ トランジション品質は、完全に主観的でも完全に客観的でもない。そこには測定可能な核がある。二つのステレオチャンネル入力とステレオマスター出力が同時に観測できるなら、粗い活動度閾値ではなく実際の寄与からトランジション区間を検出し、解釈可能な成分スコアでその区間を評価できる。

提案枠組みは、寄与認識型トランジション検出と、品質の六つの測定次元、すなわちラウドネス制御、スペクトル衝突、スペクトル連続性、軌道滑らかさ、ステレオ安定性、任意成分としてのビート整合性を組み合わせる。さらに、観測信号に対して自らのモデル説明力が低い場合には、そのことを認めるための信頼度も返す。この正直さは重要である。自信満々に間違ったスコアを返すくらいなら、スコアがない方がまだましだ。

次に必要なのはレトリックではない。データである。専門家判断付きのトランジションコーパスを構築し、成分スケールと重みをキャリブレーションし、このシステムが本当にDJの改善に役立つか検証しなければならない。もしそれが機能するなら、この分野は長く欠いていたものを得る。すなわち、センスという言葉より精密で、恣意的な単一スコアより幼稚でない、トランジション品質を議論するための真面目な言語である。

A 推奨デフォルト設定

表3は、最初の実装に対する推奨デフォルトを示す。これは普遍的真理ではない。あくまで無難に振る舞う出発値である。

表 3: 初期実装のための推奨デフォルト.

設定項目	推奨デフォルト
サンプルレート	44.1 または 48 kHz
フレーム長と hop	2048 から 4096 サンプル, hop は 512 から 1024
聴覚帯域数	およそ 30 Hz から 15 kHz を覆う 24 帯域
トランジション開始閾値	漏れ量とノイズフロアに応じて 0.15 から 0.25
τ_{on}	
トランジション終了閾値	τ_{on} より低いヒステリシスを持たせて 0.08 から 0.15
τ_{off}	
最小活性化長	$L_{\text{on}} \approx 0.5$ 秒, $L_{\text{off}} \approx 0.75$ 秒
ラウドネス許容値 δ_L	より大きい有効デッキに対して 1.5 LU 上まで
true peak 上限 P_{max}	保守運用なら -1 dBTP
低域重み ω_b	およそ 200 Hz 以下は二倍重み, およそ 1 kHz までは緩やかに強調
総合重みの初期値	表 2 に従い, その後再調整
信頼度の扱い	Conf が 0.5 などの閾値を下回る場合, 最終スコアを抑制またはグレー表示

B 参照実装の概要

参照実装は次の順序で構成できる。

1. A , B , C を共通サンプルレートへ再標本化し, フレーム境界を揃える。
2. mid と side を計算し, その後 STFT を計算する。
3. 聴覚帯域へ集約し, 式 (5) の平滑非負寄与問題を解く。
4. 寄与活動度 $\mu(m)$ から m_s と m_e を検出する。
5. T 上でラウドネス, 衝突, 連続性, 滑らかさ, ステレオ, 任意のビートスコアを計算する。
6. 式 (11) から信頼度を計算する。
7. 成分スコアと総合スコアの両方を返す。総合スコアだけを返してはならない。

参考文献

- [1] D. Cliff, “Hang the DJ: Automatic Sequencing and Seamless Mixing of Dance-Music Tracks,” HP Laboratories Technical Report HPL-2000-104, 2000.
- [2] D. Cliff, “hpDJ: An Automated DJ with Floorshow Feedback,” in *Consuming Music Together: Social and Collaborative Aspects of Music Consumption Technologies*, K. O’Hara and B. Brown, Eds., Computer Supported Cooperative Work, pp. 241–264, Springer, 2006.
- [3] H. Ishizaki, K. Hoashi, and Y. Takishima, “Full-Automatic DJ Mixing System with Optimal Tempo Adjustment based on Measurement Function of User Discomfort,” in *Proceedings of the 10th International Society for Music Information Retrieval Conference*, 2009.

- [4] T. Hirai, H. Doi, and S. Morishima, “MusicMixer: Automatic DJ System Considering Beat and Latent Topic Similarity,” in *Proceedings of the 22nd International Conference on MultiMedia Modeling*, vol. 9516 of *Lecture Notes in Computer Science*, pp. 698–709, 2016.
- [5] R. M. Bittner, M. Gu, G. Hernandez, E. J. Humphrey, T. Jehan, P. H. McCurry, and N. Montecchio, “Automatic Playlist Sequencing and Transitions,” in *Proceedings of the 18th International Society for Music Information Retrieval Conference*, pp. 442–448, 2017.
- [6] L. Vande Veire and T. De Bie, “From Raw Audio to a Seamless Mix: Creating an Automated DJ System for Drum and Bass,” *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2018, article 13, pp. 1–21, 2018.
- [7] B.-Y. Chen, W.-H. Hsu, W.-H. Liao, M. A. Martínez-Ramírez, Y. Mitsufuji, and Y.-H. Yang, “Automatic DJ Transitions with Differentiable Audio Effects and Generative Adversarial Networks,” in *2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 466–470, 2022.
- [8] A. Williams, G. Meehan, S. Lattner, J. Pauwels, and M. Barthelet, “Temporal Considerations in DJ Mix Information Retrieval and Generation,” in *32nd International Symposium on Temporal Representation and Reasoning (TIME 2025)*, vol. 355 of *Leibniz International Proceedings in Informatics (LIPIcs)*, Article 20, pp. 20:1–20:8, 2025.
- [9] R. B. Gebhardt, M. E. P. Davies, and B. U. Seeber, “Psychoacoustic Approaches for Harmonic Music Mixing,” *Applied Sciences*, vol. 6, no. 5, article 123, 2016.
- [10] D. Schwarz and D. Fourer, “Towards Extraction of Ground Truth Data from DJ Mixes,” in *Extended Abstracts for the Late-Breaking/Demo Session of the 18th International Society for Music Information Retrieval Conference*, 2017.
- [11] D. Schwarz and D. Fourer, “UnmixDB: A Dataset for DJ-Mix Information Retrieval,” in *Extended Abstracts for the Late-Breaking/Demo Session of the 19th International Society for Music Information Retrieval Conference*, 2018.
- [12] D. Schwarz and D. Fourer, “Methods and Datasets for DJ-Mix Reverse Engineering,” in *Perception, Representations, Image, Sound, Music: 14th International Symposium, CMMR 2019, Revised Selected Papers*, vol. 12631 of *Lecture Notes in Computer Science*, pp. 31–47, Springer, 2021.
- [13] T. Kim, M. Choi, E. Sacks, Y.-H. Yang, and J. Nam, “A Computational Analysis of Real-World DJ Mixes Using Mix-To-Track Subsequence Alignment,” in *Proceedings of the 21st International Society for Music Information Retrieval Conference*, pp. 764–770, 2020.
- [14] T. Kim, Y.-H. Yang, and J. Nam, “Reverse-Engineering the Transition Regions of Real-World DJ Mixes Using Subband Analysis with Convex Optimization,” in *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2021.
- [15] T. Kim, Y.-H. Yang, and J. Nam, “Joint Estimation of Fader and Equalizer Gains of DJ Mixers Using Convex Optimization,” in *Proceedings of the 25th International Conference on Digital Audio Effects (DAFx20in22)*, pp. 312–319, 2022.
- [16] É. P. André, D. Fourer, and D. Schwarz, “DJ Mix Transcription with Multi-Pass Non-Negative Matrix Factorization,” *arXiv preprint arXiv:2410.04198*, 2024.

- [17] D. Schwarz, D. A. Schindler, and S. Spadavecchia, “A Heuristic Algorithm for DJ Cue Point Estimation,” in *Proceedings of the 15th International Sound and Music Computing Conference*, pp. 259–264, 2018.
- [18] M. Zehren, M. Alunno, and P. Bientinesi, “M-DJCUE: A Manually Annotated Dataset of Cue Points,” in *Extended Abstracts for the Late-Breaking/Demo Session of the 20th International Society for Music Information Retrieval Conference*, 2019.
- [19] M. Zehren, M. Alunno, and P. Bientinesi, “Automatic Detection of Cue Points for the Emulation of DJ Mixing,” *Computer Music Journal*, vol. 46, no. 3, pp. 67–82, 2022.
- [20] G. Argüello, L. A. Lanzendörfer, and R. Wattenhofer, “Cue Point Estimation using Object Detection,” *arXiv preprint arXiv:2407.06823*, 2024.
- [21] E. Perez-Gonzalez and J. D. Reiss, “Automatic Mixing: Live Downmixing Stereo Panner,” in *Proceedings of the 10th International Conference on Digital Audio Effects (DAFx-07)*, pp. 63–68, 2007.
- [22] E. Perez-Gonzalez and J. D. Reiss, “Automatic Gain and Fader Control for Live Mixing,” in *2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 1–4, 2009.
- [23] E. Perez-Gonzalez and J. D. Reiss, “A Real-Time Semiautonomous Audio Panning System for Music Mixing,” *EURASIP Journal on Advances in Signal Processing*, Article ID 436895, 2010.
- [24] P. D. Pestana and J. D. Reiss, “Intelligent Audio Production Strategies Informed by Best Practices,” in *Proceedings of the AES 53rd International Conference: Semantic Audio*, pp. 1–9, 2014.
- [25] B. De Man, B. Leonard, R. L. King, and J. D. Reiss, “An Analysis and Evaluation of Audio Features for Multitrack Music Mixtures,” in *Proceedings of the 15th International Society for Music Information Retrieval Conference*, pp. 137–142, 2014.
- [26] B. De Man, M. Boerum, B. Leonard, R. King, G. Massenburg, and J. D. Reiss, “Perceptual Evaluation of Music Mixing Practices,” presented at the *138th Audio Engineering Society Convention*, Convention Paper 9235, 2015.
- [27] B. De Man, J. D. Reiss, and R. Stables, “Ten Years of Automatic Mixing,” in *Proceedings of the 3rd Workshop on Intelligent Music Production*, 2017.
- [28] D. Moffat and M. B. Sandler, “Approaches in Intelligent Music Production,” *Arts*, vol. 8, no. 4, article 125, 2019.
- [29] M. A. Martínez-Ramírez, D. Stoller, and D. Moffat, “A Deep Learning Approach to Intelligent Drum Mixing with the Wave-U-Net,” *Journal of the Audio Engineering Society*, vol. 69, no. 3, pp. 142–151, 2021.
- [30] M. A. Martínez-Ramírez, W.-H. Liao, G. Fabbro, S. Uhlich, C. Nagashima, and Y. Mitsufuji, “Automatic Music Mixing with Deep Learning and Out-of-Domain Data,” in *Proceedings of the 23rd International Society for Music Information Retrieval Conference*, 2022.
- [31] D. Barchiesi and J. D. Reiss, “Reverse Engineering of a Mix,” *Journal of the Audio Engineering Society*, vol. 58, no. 7/8, pp. 563–576, 2010.
- [32] M. Ramona and G. Richard, “A Simple and Efficient Fader Estimator for Broadcast Radio Unmixing,” in *Proceedings of the 14th International Conference on Digital Audio Effects (DAFx-11)*, pp. 265–268, 2011.

- [33] J. T. Colonel and J. D. Reiss, “Reverse Engineering of a Recording Mix with Differentiable Digital Signal Processing,” *Journal of the Acoustical Society of America*, vol. 150, no. 1, pp. 608–619, 2021.
- [34] S. Lee, M. A. Martínez-Ramírez, W.-H. Liao, S. Uhlich, G. Fabbro, K. Lee, and Y. Mitsufuji, “Reverse Engineering of Music Mixing Graphs with Differentiable Processors and Iterative Pruning,” *arXiv preprint arXiv:2509.15948*, 2025.
- [35] International Telecommunication Union, *Recommendation ITU-R BS.1770-5: Algorithms to Measure Audio Programme Loudness and True Peak Audio Level*, 2023.
- [36] D. P. W. Ellis, “Beat Tracking by Dynamic Programming,” *Journal of New Music Research*, vol. 36, no. 1, pp. 51–60, 2007.
- [37] S. Böck, F. Krebs, and G. Widmer, “Joint Beat and Downbeat Tracking with Recurrent Neural Networks,” in *Proceedings of the 17th International Society for Music Information Retrieval Conference*, pp. 255–261, 2016.
- [38] M. Heydari and Z. Duan, “BeatNet+: Real-Time Rhythm Analysis for Diverse Music Audio,” *Transactions of the International Society for Music Information Retrieval*, vol. 7, no. 1, pp. 274–287, 2024.
- [39] M. Torcoli, T. Kastner, and J. Herre, “Objective Measures of Perceptual Audio Quality Reviewed: An Evaluation of Their Application Domain Dependence,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 1530–1541, 2021.